Yapay Zekâ Etiğine İnsan Temelli Yaklaşımın Bir Eleştirisi

Egemen Bahadır DUMLUDAĞ[1]

Özet

Yapay zekâ alanındaki teknolojik ilerlemeler hukukun ortaya çıkmakta olan bu yeni alanda düzenlemeler getirmesi konusunda gereklilikler doğururken bu düzenlemeler yapılırken etik de bir kaynak olarak karşımıza çıkacaktır. Yapay zekâ etiğine insan temelli yaklaşım literatürde, yapay zekâ etiği üzerine etik düşüncenin merkezinde insan refahı ve sorumluluğuna vurgu yapan ve henüz yapay zekânın etik olarak sorumlu olamayacağı düşüncesini temel alan bir yaklaşımdır. İşbu çalışma kapsamında (1) bu yaklaşıma karşı, hem yapay zekânın ve insanın etik düşünce şeklindeki benzerlikler ve yaklaşım içerisindeki bazı tutarsızlıkları ortaya koyarak bir eleştiri getirmeye çalışılacak, (2) yapay zekâ kullanan araçların ne zaman araç, ne zaman etik süje yahut hukukî sorumlu olarak değerlendirilebileceği tartışılacak, (3) etik süje rolünün hukukî sorumluluğa muhtemel etkisine değinilecek, ve (4) Julia Kristeva'nın *abjection* kavramı üzerinden eleştirilen makaledeki tutarsızlıkların nedeni sorgulanacaktır.

Anahtar Sözcükler: yapay zekâ, etik, sorumluluk, *abjection*

Yapay zekâ kavramı yeni bir kavram olmamakla birlikte yıllar içerisinde farklı tanımları kapsar bir şekilde kendisine literatürde yer buldu. Buna karşın, son zamanlarda özellikle makine öğrenmesi alanındaki gelişmelerle birlikte yapay zekâ ve makine öğrenmesi kavramları neredeyse birbirlerinin yerine kullanılır durumdalar. Bu güncel gelişmeler ışığında yapay zekâ konusu literatürde ve yasama çalışmalarında daha yoğun "kaygı konusu" haline gelmiş durumda.

Yapay zekâ konusundaki bu ilerlemelerle birlikte gündeme gelen sorulardan birisi yapay zekânın etik süje olarak rolü, bu etik alanında soyut bir tartışma olmaktan öte hukukî sorumluluğa da etki edebilecek somut bir sorunu temsil ediyor. Bu kapsamda ortaya atılan fikirlerden birisi ise "yapay zekâ etiğine insan temelli yaklaşım". Bu yaklaşımın temel tezi yapay zekâ sistemlerinin henüz etik süje rolünü haiz olmadıkları ve etik sorumluluğun incelenmesinde insan odaklı hareket edilmesi gerekliliği.

1 Ankara Üniversitesi Hukuk Fakültesi, Hukuk Felsefesi ve Sosyolojisi Anabilim Dalı Yüksek Lisans Öğrencisi
E-posta: egemen1109@gmail.com

The Oxford Handbook of Ethics of AI kitabında kendisine yer bulan "insan temelli yaklaşım" tezi yukarıda bahsedilen araştırma kaygısının bir ürünü olarak yapay zekânın etik süje olarak bir rolü olamayacağını iddia ederken zaman zaman çeşitli tutarsızlıklar içeren iddialar öne sürüyor. Bu makalede ilgili kitapta yer alan "İnsan Temelli Yaklaşım" makalesi eleştiriye konu edilecek ve bu kapsamda, (1) ilgili makalenin içerdiği tutarsızlıklar ortaya koyularak öne sürülen tez, (2) yapay zekâ sistemlerinin ne zaman sorumlu ve ne zaman yalnızca birer araç olarak nitelenebileceği, ve (3) etik süje rolünün hukukî sorumluluğa etkisi tartışılacak, bunlardan sonra ise kısaca (4) Julia Kristeva'nın *abjection* kavramı üzerinden yapay zekânın etik sorumluluğuna ilişkin bu yaklaşımların nedeni sorgulanacaktır.

"A Human-Centered Approach to AI Ethics" makalesinde anlatıldığı üzere yapay zekâ etiğine insan temelli yaklaşımın iki temel ögesi mevcuttur: (1) insan refahına vurgu, (2) insan sorumluluğuna vurgu[2]. Makaleye göre insan refahına vurgu "Robotlar acıyı hissedebilirler mi?", "Robotlar acı çekebilirler mi?" gibi sorularla mücadele eden yaklaşımlara ters düşer[3]. Benzer şekilde insan sorumluluğuna vurgu ise "Mücrim robotları cezalandırabilir miyiz?", "Bazı makinelere vatandaşlık, işçi hakları, 'insan' hakları tanıyabilir miyiz?" gibi sorularla ilgilenen yaklaşımlarla ters düşer[4]. Yazara göre insan temelli yaklaşım bu sorulara yalnızca "hayır" cevabı vermekle kalmaz, aynı zamanda onları "küstah" olarak nitelendirir, hatta gerçek etik sorunlardan bizi uzaklaştıran dikkat dağıtıcılar olarak[5].

Yaklaşım, bu iki vurgunun yanında iddialarına deflasyonist görüş ile başlıyor, makaleye göre bu görüş şu şekilde tanımlanıyor: Güncel olarak var olan hiçbir robot yahut yapay zekâ (YZ) sistemi etik olarak sorumlu olamaz, yahut kendisine yönelik etik sorumluluğumuz olacak neviden bir şey olamaz[6]. Yazarın kendisi de bunu argümandan yoksun bir şekilde ortaya attığını kabul etmektedir ve buna rağmen deflasyonist görüşün daha güncel ve somut sorunlara odaklanma yolunu açtığını savunmaktadır[7]. Bu kapsamda "Etik

2 Ron CHRISLEY, "A Human-Centered Approach to AI Ethics: A Perspective from Cognitive Science", *The Oxford Handbook of Ethics of AI* içinde, Derleyen: Markus D. DUBBER *et al.*, Oxford UP, New York, 2020, s. 464.

3 CHRISLEY, 2020, s. 464.

4 CHRISLEY, 2020, s. 465.

5 CHRISLEY, 2020, s. 465.

6 CHRISLEY, 2020, s. 465.

7 CHRISLEY, 2020, s. 465.

robotlar yapmaya çalışmayın, robotları etik olarak yapmaya çalışın." sloganını benimsediğini söylemektedir[8].

Yukarıda temelleri özetlenmeye çalışılan fikirlerle yola çıkan metnin kendisine başlangıç noktası olarak aldığı deflasyonist görüş onun eleştirisinin de en önemli sütunlarından. John P. Sullins'in "When Is a Robot a Moral Agent?" makalesinde yapay zekâ sistemlerinin birer *agent* olabilmesi için gerekecek koşullara ilişkin dört farklı görüşe yer veriliyor, yazar bu makaleyi de örnek vererek deflasyonist görüşü kapsamında bu şartların hiçbirinin öngörülebilir gelecekte gerçekleşmeyeceğini düşündüğünü beyan ediyor[9].

Sullins dört farklı görüşle birlikte üç sorunun önemli olduğunu belirtiyor[10]:

- Robot önemli ölçüde otonom mu?
- Robotun davranışları kastî mi?
- Robot sorumlu olduğu bir pozisyonda mı?

Sullins için bunlar daha önceden belirttiği dört görüşten sonuncusu olan pragmatik soyutlama düzeylerinden uygun olan biri kapsamında incelenmeli. Pragmatik soyutlanma düzeyleri önemli bir eleştiri noktasına parmak basıyor: yeterince düşük bir soyutlanma düzeyinde insanlar da mekanizmadır, ancak doğru soyutlanma düzeyleri pragmatik olarak anlamlı incelemeler sağlayacaktır[11].

Yazar Sullins'e atıf verip bu şartlar her ne kadar doğru olabilirse de bunlar yakın zamanda gerçekleşmemiş olacaktır demekle Sullins'in üç sorusuna birden "evet" diyemeyeceğimiz iddiasını ortaya atmış oluyor.

Otonomi sorusuyla ilgili Sullins'in düşüncelerini özetlemek gerekirse ilk gündeme gelen, bir robotun programcılarından, operatörlerinden ve kullanıcılarından yeterince bağımsız kabul edilip edilemeyeceği sorusu[12]. Sullins'e göre felsefî özerklik kavramı karmaşık olsa da, yazar özerkliği "teknik" anlamında, yani robotun herhangi bir haricî aracı veya kullanıcı tarafından doğrudan kontrol edilmediği anlamında kullanıyor[13].

8 CHRISLEY, 2020, s. 473.

9 CHRISLEY, 2020, s. 465^3.

10 John P. SULLINS, "When Is a Robot a Moral Agent?", *International Review of Information Ethics*, Vol. 6, No. 12, 2006, s. 28.

11 SULLINS, 2006, s. 28.

12 SULLINS, 2006, s. 28.

13 SULLINS, 2006, s. 28.

Kasıtlılık, yahut niyetlilik sorusu için ise Sullins davranışın halk nezdindeki kötülük veya iyiliğe eğilim yahut "niyet" gibi psikolojik kavramlara başvurularak açıklanması gerekiyorsa soruya olumlu yanıt verilebileceğini söylüyor[14]. Ona göre tek gereken *agent*ların etkileşimleri düzeyinde kıyaslanabilir ölçüde kişisel bir niyetlilik ve özgür irade olması[15].

Sullins'in son sorusu olan "sorumluluk" ise meselenin karmaşıklaştığı ve eleştirilerin doğduğu bir nokta. Yine de bu çalışma Sullins'in makalesindeki şartları eleştirmeye genişlemiyor, yalnızca insan temelli yaklaşım içerisindeki tutarsızlıkları gözlemleme çabasını kapsıyor. Sullins'e göre son şart, robot diğer *moral agent*lara karşı bir sorumluluğu olduğu şekilde davrandığında gerçekleşmiş oluyor[16]. Eğer robot bu şekilde davranıyorsa ve bazı verilmiş sorumluluklarla bir sosyal rolü üstleniyorsa ve bu davranışı anlamamızın tek yolu bu hastalarla ilgilenmesi görevine sahip olduğu inancı ise o halde bu makinenin bir *moral agent* olduğu sonucuna varılabilir[17]. Sullins bu inancın gerçek bir inanç olması gerekmediğini sadece görünürde olması gerektiğini vurguluyor; makinenin bir bilinç, örneğin bir ruh, zihin ya da insanın özel olduğuna sebep kabul edilir şeylerin hiçbirine iddiası olmayabilir, bu inançlar sadece makinenin ahlâki çıkmazlarında ona yol göstermelidir[18].

Bir örnek olarak yaşlıların bakımını üstlenen robotik bakıcılar ele alındığında Sullins'e göre bir hemşire kesinlikle bir *moral agent* iken bir makinenin aynı statüyü elde etmesi için yukarıdaki gibi otonom, niyetli hareket eden ve kendisinin yer aldığı sağlık sisteminin bakımla yükümlü olduğu hastalara karşı rolünü anlayan bir sistem olması gerekir[19].

Hem insan temelli yaklaşım hem Sullins için tutarsızlık noktası bu üç soruda da mevcut. Mevcut YZ sistemleri için ilk soru kapsamındaki otonomluğun sağlanması oldukça basit, YZ sistemleri bir operatör olmadan karar alma kabiliyetine sahipler. Kaldı ki bu otonomluk çok daha basit sistemler için bile mümkün: bir evdeki duman sensörü bile bir kullanıcıdan bağımsız olarak duman tespit ettiğinde bağırmaya mukabil. Hatta güncel LLM (*large language model*, *büyük dil modeli*) modellerinin bir kısmında sorun belki

14 SULLINS, 2006, s. 28.

15 SULLINS, 2006, s. 28.

16 SULLINS, 2006, s. 28.

17 SULLINS, 2006, s. 28.

18 SULLINS, 2006, s. 29.

19 SULLINS, 2006, s. 29.

de onları otonom kılmaktan değil onları daha az otonom kılmaktan doğuyor.

Kasıtlılık sorusu biraz çetrefilli; zira Sullins'in kendisi çok detaylı açıklamadığı gibi güçlü bir şekilde ispatı gerekmediğini de savunuyor. Sullins'e göre makinenin davranışları çevresiyle uyumlu, ahlâki olarak zararlı yahut faydalı ve niyetli ve hesaplanmış görülüyorsa makine niyetli değerlendirilebilir[20]. Bu noktada Sullins'in tartışmaya davet ettiği pragmatik soyutlanma düzeylerini gündeme getirmek mümkün. İnsanlar için de birçok davranış psikolojik ceza ve ödül kavramlarıyla ilişkili, toplum nezdinde makbul vatandaş olmanın faydalarının çekiciliği gibi. Kaldı ki Sullins de sorunun insanlar için de zor olduğunu söylüyordu. O halde genetik algoritmalar ile tasarlanmış ve çevresinde insan öldüğünde negatif ödül alan bir satranç robotunun satrançta hesaplayarak hamle yapması onun için bu şartı sağlayacaktır, yani mevcut kabiliyetler dahilinde bu şartı da sağlamak mümkün görünüyor.

Son soru için ise Sullins'in yaşlı hastalara bakan robot örneği üzerinden gitmek mümkün. Sullins robotun sağlık sisteminin rolünü anlaması gerektiğini öne sürüyor. Açıkçası "anlamak" ile neyi kastettiğini anlamak mümkün değil. Bir insanın bu sağlık sisteminin rolüne yönelik anlayışını nasıl ifade edebileceğimiz de tartışılır. Anlamak epistemolojik yük üstlenen bir eylem ise "bilgi nedir?" sorusunun bile tartışıldığı bir literatürde böyle bir cümle ortaya atmak koca bir belirsizlik kapısı aralıyor.

"Anlama"ya yönelik anlama zorluklarına rağmen sağlık sisteminin rolünün insanları iyileştirmek olduğu kabulüyle hareket edersek, ki buna dahi siyaset bilimi çerçevesinden eleştiriler getirmek mümkün olacaktır, robotun tek "bilmesi" gereken hastaların iyileşmesinin gerektiği. "İyileşmek" kelimesini bir robota anlatmak oldukça güç bir sorun gibi görünebilir, hatta yapay zekâ literatürünün gölgede kalmış güvenlik alanındaki "alignment" problemleri de buna benzer sorunlarla ilgilenmektedir. Nasıl ki insanlar zaman zaman "iyi olmak" ne demek sorgulamaktaysa, bu kavramları bir sisteme anlatmak da aynı şekilde güç. Yine de elimizdeki teknoloji ile bir sürü sağlıklı insanı bir sisteme gösterip olması gerekene dair bir fikir elde etmesini sağlayabiliriz, zira bu aslında bir köpek ile bir kediyi ayırt etmekten pek farklı bir sorun değil, günün sonunda bir iki resim arasındaki farkları bulma oyunu. Bunu yeterli bulmadığımız noktada bir insanın da sağlık sisteminin rolünü anlamaktaki güçlükleri yine Sullins ve insan temelli yaklaşıma karşı tezler olarak gündeme geliyor.

20 SULLINS, 2006, s. 28.

Deflasyonist görüşle birlikte YZ'nin sorumlu olması göz ardı edilirken insan temelli yaklaşım aynı zamanda YZ'nin etik bakımdan kaygı konusu olması sorusunu da küstah bulunuyordu. Oysa makalenin hayır cevabını vermekle yetinmediği "Robotlar acı hissedebilir mi?" sorusu küstah nitelemesinden öte, "evet" yanıtını da haiz olabilir. Makalenin bu konuda sunduğu hiçbir net iddia olmamakla birlikte acının mekanizmasının makalede değerlendirilmediği görülebilir. Öyle ki "acı"nın tanımında eğer bilişsellikten bağımsız tanımlar kullanılırsa YZ sistemlerinin bir kısmının geliştirilmesinde kullanılan negatif ödül sistemlerinin primitif acı mekanizmalarına benzediği tartışılabilir.

"Evolution of Mechanisms and Behaviour Important for Pain" makalesinde "acının en nüfuzlu tanımı kendimizinki dışındaki tüm türleri göz ardı eder" şeklinde ifade edilen sorun insan temelli yaklaşımı da sarmış durumdadır[21]. Buna çözüm olarak Robert W. Elwood, acıyı öznel yerine fonksiyonel özellikleri üzerinden tanımlamayı önermektedir[22]. Durum bu olduğunda yukarıda belirtildiği şekilde negatif ödül sistemleri istenilen sonuçların alınması adına konulmuş öğrenme mekanizmaları olarak acının evrimsel rolüne benzer bir rol üstlenmektedir; zira Elwood'un da belirttiği üzere acı hayatta kalma ihtimalini güçlendiren bir fonksiyondadır[23]. O halde robotların acı çekemeyeceği ifadesi havada bırakılmıştır.

Makale kendi tezlerini havada bırakarak ilerleyerek insan temelli yaklaşımın uygulanabileceği örneklere geçerken bir askeri robot örneği üzerinde yoğunlaşıyor: Bu R robotu, A ve B köprülerinin olduğu bir savaş bölgesinde H insanı kontrolü altında çalışıyor[24]. R, A ve B köprülerinin olduğu bölgede devriye gezmesi için H tarafından harekete geçirilebiliyor ve R'nin kabiliyetleri arasında köprüyü patlatmak da mevcut[25]. Yazara göre bu durumda köprüyü yıkmak etik olarak iyi; zira bu masumları saldırıdan koruyacak; fakat köprünün üzerinde medikal malzemelerin de olduğu mini hastaneler olduğu durumlar hariç[26]. Buna uygun şekilde R, H ile irtibata geçemediği durumlarda dahi üzerinde

21 Edgar T. WALTERS ve Amanda C. de C. WILLIAMS, "Evolution of Mechanisms and Behaviour Important for Pain", *Phil. Trans. R. Soc. B*, 2019, 374:20190275, s. 2.

22 WALTERS ve WILLIAMS, 2019, s. 2.

23 Robert W. ELWOOD, "Discrimination Between Nociceptive Reflexes and More Complex Responses Consistent With Pain in Crustaceans", *Phil. Trans. R. Soc. B*, 2019, 374:20190368, s. 3.

24 CHRISLEY, 2020, s. 471.

25 CHRISLEY, 2020, s. 471.

26 CHRISLEY, 2020, s. 471.

hastane yoksa köprüleri patlatacak şekilde tasarlanmış[27].

H'nin R'yi harekete geçirdiği sırada H büyük olasılıkla B üzerinde bir hastane olduğuna, ancak A üzerinde bir hastane olmadığına inanıyor[28]. A'ya giden yol sırasında B'nin yanından geçen R, kameralarıyla B üzerinde bir hastane olmadığı kanısına varıyor ve B'yi patlatıyor[29]. Ancak, R'nin edinmiş olduğu bilgi yanlış ve B üzerinde bir hastane mevcut[30]. Makale, bu süreç üzerinden insan temelli yaklaşımın pratiğe dökülebilmesi için hem çıkarımsal hem de robot tasarımına yönelik iki çözüm öneriyor[31]. Oysa buradaki temel sorun şu ki çalışan robot bir yapay zekâ sistemi olmakla birlikte karar alan bir yapay zekâ sistemi değil. Yapay zekâya yönelik dört temel görüş "insanlar gibi düşünen sistemler", "rasyonel düşünen sistemler", "insanlar gibi davranan sistemler", ve "rasyonel davranan sistemler" olarak basitleştirilebilirler[32]. Bu tanımlardan insanlar gibi davranan sistemler klasik Turing testi gibi ölçeklerle incelenebilirse de bu tanım yapay zekâ kavramını oldukça muğlak kılıyor. İnsanlar hesap yapabiliyorsa insanlar gibi hesap yapabilen bir hesap makinesine yapay zekâ demek pragmatik bir tanımlama olmayacaktır. Yalnızca bir "eğer" sorusuna göre hareket eden bir sisteme etik süje demek zor olacağı gibi yapay zekâ demek de zor olacaktır.

Tanımsal farklılıklara karşın, ki bu dört temel görüş dışında yapay zekâ kavramına yönelik çok daha fazla tanım bulmak elbet mümkün olacaktır, robot bakımından bir yapay zekâ nitelemesi yapılabilirse o da hastanenin varlığının tespiti nezdinde olabilecektir. Tabiî, eğer insan temelli yaklaşım transistörlere ve hatta hayvan hücrelerinin hücre duvarına zekâ deme iddiasında değilse; zira köprünün patlatılması işi basit bir "eğer" sorusuna dayalıdır: "eğer hastane varsa/yoksa". H'nin sorumluluğunun tespiti de rus ruleti benzeri bir biçimdedir. Robotun hastaneyi yanlış tespit ettiği olasılığını gözeterek hareket eden H, bu olasılığın şiddetine göre "olası kast" gibi kurumlarla değerlendirilecektir; ancak epistemolojik olarak H'nin elindeki tek veri oynadığı kumardaki yanlış pozitif ve yanlış negatif olasılıklarıdır. Oysa deflasyonist görüşle yok sayılan robotik beceriler daha karmaşık karar alma mekanizmalarına da izin verir niteliktedir. Otonom, kastî hareket eden,

27 CHRISLEY, 2020, s. 471.

28 CHRISLEY, 2020, s. 472.

29 CHRISLEY, 2020, s. 472.

30 CHRISLEY, 2020, s. 472.

31 CHRISLEY, 2020, s. 472.

32 Stuart J. RUSSELL ve Peter NORVIG, *Artificial Intelligence: A Modern Approach*, Prentice Hall, New Jersey, 1995, s. 5.

ve sorumlu olan bir robot tasarlamak ve buna karar aldırmak mümkün, en azından teknik olarak mümkün, öyleyse deflasyonist görüşü ayakta tutmak mümkün değil. Bu, insan temelli yaklaşımın terk edilmesi gerektiğine işaret etmiyor, yahut gerçekten yapay zekâ ile çalışan robotlara ihtiyacımız olduğuna. Yapılabilecek tek çıkarım insan temelli yaklaşımın temellerinin sağlam olmadığı, doğruluğu ya da yanlışlığını tartışmak kapsam dışı olacaktır.

Yapay zekâ sistemlerinin etik sorumluluğunun tartışılmasının hukuk bakımından önemi ise etik sorumluluğun hukukî sorumluluğa etkisi bağlamında doğuyor. Hart, "Positivism and the Separation of Law and Morals" makalesinde ahlâki kurallar ve prensiplerin hukuka etkisini faydacıların kabulünden şu şekilde bahsediyor:

"Öncelikle, onlar asla tarihsel bir gerçek olarak hukukî sistemlerin ahlâkî görüşlerden etkilendiğini ve tersi olarak ahlâkî standartların hukuktan çokça etkilendiğini ve böylece çoğu hukuk kuralının ahlâkî kural ve prensiplerin aynası olduğunu inkâr etmediler."[33]

Hatta Alberto Bondolfi ve Jason Nye bundan bir adım daha ileri giderek etikçilerin yasamadaki rolünü de tartışma konusu yapıyor[34].

Güncel hukuk düzleminde yasaların gerekçelerinin dönemin algısına göre değiştiğini gözlemlemek mümkündür. 12 Mayıs 2003 tarihli ve B.02.0.KKG.0.10/101-540/2092 sayılı Türk Ceza Kanunu tasarısının gerekçesinde de bu gözlemlenebilir. Tasarının 139. maddesi "günümüzde bu konuda XIX. yüzyılın görüşleri değişmiştir" diyerek bunu ortaya koymaktadır.

Özellikle bir felsefe alanı olarak etik tartışılırken etik ve ahlâkın ayrımının yapılması gerekecektir. Etik bir felsefî bilgi alanı iken, ahlâk belirli bir insanın karşılaştığı belirli bir durumda ne yapılması gerektiğiyle ilgilenir. Yine de burada farkında olmak gerekir ki "Doğru eylem nedir?" yahut "Yapay zekâ etik süje midir?" gibi etiğe dayalı felsefî sorular ahlâkî tartışmalara da etki edecek ve böylece hukukun da alanına sızmaya imkân bulacaktır. Kaldı ki, konuyla ilgili tartışmalarda Joanna Y. Bryson gibi bazı yazarlar etik tartışmada YZ'nin süje olarak rolünü, ahlâkî süjeyi toplum tarafından sorumlu tutulan şeyler olarak tanımlayarak yalnızca toplumsal olarak değerlendirmeye de yönelmişlerdir.[35]

Yapay zekânın ilerlemesiyle hukukî sorumluluk tartışmaları da onu takip eder hâlde. Avrupa Konseyinin

33 H. L. A. HART, "Positivism and the Seperation of Law and Morals", *Harvard Law Review*, Vol. 71, No. 4, Şub. 1958, s. 598.

34 Alberto BONDOLFI ve Jason NYE, "Ethics, Law and Legislation: The Institutionalisation of Moral Reflection", *Ethical Theory and Moral Practice*, Vol. 3, (Mar. 2000), No. 1, Justice in Philosophy and Social Science, s. 32.

35 Joanna J. BRYSON, "Patiency is not a Virtue: The Design of Intelligent Systems and Systems of Ethics", *Ethics and Information Technology*, Vol. 20, 2018, s. 16.

"A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework" isimli çalışması da bunlardan biri. Çalışma kapsamında YZ'nin özellikle hangi açılardan sorumluluk alanlarına etkili olduğu tartışılırken özellikle insan hakları perspektifine yoğunlaşılıyor. Çalışma YZ'nin "sorumluluk bakımından önemli" özelliklerini, diğerlerini etkileyebilecek özellikleri olarak tanımlıyor[36]. Bu bakımdan tartışmanın henüz YZ'nin sorumluluğunu tartışma noktasına gelmediğini ve insan temelli yaklaşım ile paralel olduğunu söylemek mümkün.

Makalenin "logic-based ethical robot methodology" başlığı altında savunduğu, robotların yine makalenin ifadesiyle "explicit moral agent" olması[37] hukukî sorumluluk noktasında başka tartışmaların da önünü açıyor. Hukukî sorumluluk için ne tür bir *agency* gerektiği sorusunun da bu tartışmada yeri vardır[38]. Kaldı ki hukukî sistemlerde önemli kabul edilebilecek dört *agency*den biri de bu "explicit moral agent" denilen *agency* olarak karşımıza çıkmaktadır: (1) açık ahlâkî *agent*lık, (2) örtülü ahlâkî *agent*lık, (3) açık epistemik *agent*lık, ve (4) örtük epistemik *agent*lık[39].

Tüm bu tartışmalar sonrasında yapay zekâ için kesin bir sorumluluk atfı, yahut ona etik bir süje statüsü vermek bu çalışmanın kapsamının dışında kaldığından bu tartışmalar bir noktada sonuçsuz görünmektedir. Açıkçası amaçlanan da bir noktada budur. Yeni bir "şey" için etik süje rolünün insan temelli yaklaşımda olduğu gibi tutarsız argümanlarla ve bazı noktalar tartışılmadan bile bir kenara atılması etik bilgi alanını baltalamaktadır. Bu da bu çalışmayı sonucuna ve son yanıtlamaya çalıştığı soruya götürmektedir, "neden" sorusuna.

Yapay zekâya yaklaşıma belki de bu teknolojinin adıyla başlamak gerekir. Yapay zekâ bir zekâyı çağrıştırmaktadır. İnsanın kendisine *homo sapiens* adını verişindeki kendi biricikliğini arayışına da bir tehdittir bu neredeyse. İnsan zeki olmakla kalmaz, insan aslında zeki insandır, diğer insanlardan da ayrı olarak. Yapay zekâ onun gibidir ve ondandır. Doğan reaksiyonları da belki de bu bağlamda incelemek gerekir.

36 Karen YEUNG, *A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility Within a Human Rights Framework*, DGI(2019)05, Https://Rm.Coe.İnt/A-Study-Of-The-İmplications-Of-Advanced-Digital-Technologies-İncluding/168096bdab, 17.03.2023, s. 18.

37 CHRISLEY, 2020, s. 469.

38 Carlos MONTEMAYOR ve Enrique CÁCERES, "Agency And Legal Responsibility: Epistemic and Moral Considerations", *Problema*, Num. 13, 2019, s. 102.

39 MONTEMAYOR ve CÁCERES, 2019, s. 113.

Kristeva'nın *abject* kavramı içeriden ya da dışarıdan gelen fahiş bir tehdide yönelik bir histir, mümkün olanın, düşünülebilirin dışına atılmış anlamındadır[40]. *Abject* fırlatılmış obje olarak kişiyi anlamın çöktüğü yere çeker, bu yönüyle kişinin karşı durarak anlam kazandığı nesneden ayrılır[41]. *Abject*, bir hiç değildir yahut "ben" değildir, bir şeydir fakat onun bir şey olduğunu tanımayız[42]. Kristeva için yemeklere karşı duyulan iğrenmeyle gelen kusma *abjection* için çok temel bir örnek[43]. Yemek kişinin kendisi olmadığı halde kişinin arzusu olarak dışarı atıldığında da kişi kendisini sürgün etmiş olur[44]. Tartışmanın kapsamında *abject* Kristeva'nın temel anlatısında olduğu üzere bireyin kendisinden attığı ve karşı da durmadığı, tanımadığı "şey"i ifade eder.

Yapay zekânın etik sorumlu olarak kabul edilmeyişi için bir tez, onun zeki insan için bir *abject* öge olması olabilir. Zeki insan için yapay zekâ onun biricikliğine bir tehdittir, hatta onun toplum içindeki rolüne de müdahil olan ve kısıtlı kaynaklar bakımından da ona rekabet yaratan, soyut olmanın yanında somut bir tehdittir. YZ, insandan ayrı olduğu gibi aynı zamanda yemek gibi insanın arzu nesnesidir ve insanın kendi yaratısı ve benzeri olarak ondan çok derin bir parçanın zuhuru olarak vücut bulur. İnsan merkezli YZ etiği tartışmasının YZ'ye yönelik bazı sorulara "hayır" yanıtı vermekten öte bu soruları küstah kabul etmesi YZ'ye yönelik bu *abjection* hissinin bir tezahürü olabilir; zira bu değerlendirmeyiş onu bir "şey" olarak tanımaktan kaçınmak ve onu karşı durulan bir nesne olarak tanımlamadan ona karşı bir savunma duvarı örmektir. İnsan kendisini bilebilmek için, kendisinden olanı kendi dışına iterek bir sınır çizmeye çalışmaktadır.

"Füze, her an nerede olduğunu bilir. Bunu bilir çünkü nerede olmadığını bilir[45]." İnsan için de yapay zekânın etik rolünü anlama çabası bir yandan bugüne kadar kısıtlı şekilde elinde olmuş bir fırsatın tekrar değerlendirilmesidir. İnsan kendi etik sorumluluğunu da kendi etik sorumluluğunun ne olmadığını bilerek anlayabilir. Bu fırsatın değerlendirilmesi önemli olsa da YZ'ye yönelik eleştirilerin gerek *abjection* temelli, gerekse teknik anlamda YZ konusunda bilgi eksikliği sebebiyle zaman zaman temelsiz yahut tutarsız

40 Julia KRISTEVA, *Powers of Horror: An Essay on Abjection*, Çeviren: Leon S. ROUDIEZ, Columbia UP, New York, 1982, s. 1.

41 KRISTEVA, 1982, s. 2.

42 KRISTEVA, 1982, s. 2.

43 KRISTEVA, 1982, s. 2.

44 KRISTEVA, 1982, s. 3.

45 George GRILL, "GLCM Guidance System", *Association of Air Force Missileers Newsletter*, Vol. 5, No. 4, Ara. 1997, s. 5.

olması sonucu günün sonunda insanın kendisine yönelik bilgisinin de kirlenmesine neden olabilir. İnsan temelli yaklaşım pragmatik amaçlarla çıktığı yolda "etik robotlar yapmaya çalışmayın, robotları etik yapmaya çalışın"[46] sloganında haklı olsa bile gerekçelendirmelerindeki hatalarla etik robotların nasıl yapılacağına dair sorunlara verdiği yanıtlarda eksik kalıyor. Kaldı ki bu ikisi birbirini dışlamak zorunda değil, etik robotları etik şekilde yapmak da mümkün olsa gerek.

Güncel yapay zekâ sistemlerinin kabiliyetleri gün geçtikçe artıyor, yani eleştiriye konu deflasyonist görüş adına bu "*abjection!"* itirazının *overruled* denilerek ertelendiği bir durumda bile çok yakın vakitlerde yapay zekâ için etik sorumluluk tartışmalarının tekrar açılması olası. İnsan temelli yaklaşımın karar alma süreçlerinde robotların rolünü tam olarak anlayamaması ise bir başka mesele. Yapay zekâ sistemleri halihazırda çoğunlukla karar alabilecek noktalara konumlandırılmıyorlar; ancak bu, konumlandırılmaları önünde teknolojik bir engel olduğu anlamına gelmiyor. Avrupa Komisyonunun 21 Nisan 2021 tarihli COM/2021/206 sayılı düzenleme önerisi "insan gözetimi" vurgusunu çok sık yaparak insandan özerk bir şekilde YZ kullanımını regüle etme kaygısını yansıtıyor. YZ'nin kara kutu yapısı içerisinde üreten ve işletenlerin sorumlu tutulabilmesi için gözetim yükümlülükleri somut önem taşısa da etik sorumluluk düzleminde YZ'nin karar alabildiği durumlar daha iyi incelenmeli.

Eleştirilen makalede, natüralizmin kendimizi abartılmış robotlar olarak görmemize sebep olabileceği dile getirilmekte, bunun da ya etik *agency*nin robotlara genişletilmesine ya da hem insanlar hem robotlar için bir etik nihilizme sebep olabileceğine yönelik düşünceden bahsedildikten sonra, yazar tarafından bunun karşısında görüş alınmaktadır. Oysa insanın abartılmış bir robot olmasından anlamsız bir kaçınmaya gerek yoktur. İnsanın ahlâkî *agent* olması da Luciano Floridi ve J.W. Sanders'ın bahsettiği üzere belirli bir soyutlanma düzeyinde mümkündür sadece[47] ve derine indikçe amino asitler ve proteinlerin kimyasal reaksiyonlarından ibaret devasa bir fabrika olan insan vücudu sözcüğün tam anlamıyla bir "*robota*" olarak varlığını sürdürür, gerek Sisyphusyan bir yönde gerekse somut olarak hayatta kalmak için çabalayan bir evrim kevgirinin artığı et yığını olarak; fakat asıl mesele insanın abartılmış bir robot olup olmadığı değil, asıl mesele bu sorulardan kaçınmanın pragmatik sonuçlarının çekiciliğine kapılmanın tehlikesi: insanın ne olduğunu bilmeye yönelik şansını kaybetmesi.

46 CHRISLEY, 2020, s. 473.

47 Luciano FLORIDI ve J. W. SANDERS, "On the Morality of Artifical Agents", *Minds and Machine*, Vol. 14, 2004, s. 357.

KAYNAKÇA

- BONDOLFI, Alberto ve Jason NYE, "Ethics, Law and Legislation: The Institutionalisation of Moral Reflection", *Ethical Theory and Moral Practice*, Vol. 3, No.1, Mar. 2000, Justice in Philosophy and Social Science, https://www.jstor.org/stable/27504117, Erişim Tarihi: 17.03.2023.
- BRYSON, Joanna J., "Patiency is not a Virtue: The Design of Intelligent Systems and Systems of Ethics", *Ethics and Information Technology*, Vol. 20, 2018, https://doi.org/10.1007/s10676-018-9448-6.
- CHRISLEY, Ron, "A Human-Centered Approach to AI Ethics: A Perspective from Cognitive Science", *The Oxford Handbook of Ethics of AI* içinde, Derleyen: Markus D. DUBBER *et al.*, Oxford UP, New York, 2020.
- ELWOOD, Robert W., "Discrimination Between Nociceptive Reflexes and More Complex Responses Consistent With Pain in Crustaceans", *Phil. Trans. R. Soc. B*, 2019, 374:20190368, http://dx.doi.org/10.1098/rstb.2019.0368.
- FLORIDI, Luciano ve J. W. SANDERS, "On the Morality of Artifical Agents", *Minds and Machine*, Vol. 14, 2004, https://doi.org/10.1023/b:mind.0000035461.63578.9d.
- GRILL, George, "GLCM Guidance System", *Association of Air Force Missileers Newsletter*, Vol. 5, No. 4, Ara. 1997, https://www.afmissileers.org/newsletters/NL1997/Dec97.pdf, Erişim Tarihi: 17.03.2023.
- HART, H. L. A., "Positivism and the Seperation of Law and Morals", *Harvard Law Review*, Vol. 71, No. 4, Şub. 1958, https://doi.org/10.2307/1338225.
- KRISTEVA, Julia, *Powers of Horror: An Essay on Abjection*, Çeviren: Leon S. ROUDIEZ, Columbia UP, New York, 1982.
- MONTEMAYOR, Carlos ve Enrique CÁCERES, "Agency And Legal Responsibility: Epistemic and Moral Considerations", *Problema*, 2019, Num. 13, https://www.scielo.org.mx/scielo.php?pid=S2007-43872019000100099&script=sci_arttext&tlng=en, Erişim Tarihi: 17.03.2023.
- RUSSELL, Stuart J. ve Peter NORVIG, *Artificial Intelligence: AModern Approach*, Prentice Hall, New Jersey, 1995.
- SULLINS, John P., "When Is a Robot a Moral Agent?", *International Review of Information Ethics*, Vol. 6, No. 12, 2006, https://doi.org/10.29173/irie136.
- WALTERS, Edgar T., ve Amanda C. de C. WILLIAMS, "Evolution of Mechanisms and Behaviour Important for Pain", *Phil. Trans. R. Soc. B*, 2019, 374:20190275, http://dx.doi.org/10.1098/rstb.2019.0275.
- YEUNG, Karen, "A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility Within a Human Rights Framework", *DGI(2019)05*, https://rm.coe.int/a-study-of-the-implications-of-advanced-digital-technologies-including/168096bdab, Erişim Tarihi: 17.03.2023.